DOCUMENT RESUME

ED 229 451 TM 830 362

AUTHOR Wilburn, Kenneth T.; Felps, Barry C.
TITLE Do Pupil Grading Methods Effect Middle School Students' Achievement: A Comparison of Criterion-Referenced Versus Norm-Referenced Evaluation.
PUB DATE [Apr 83]
NOTE 25p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Criterion Referenced Tests; Elementary Education; Elementary School Mathematics; *Evaluation Methods; *Mathematics Achievement; Measurement Techniques; Middle Schools; *Norm Referenced Tests; *Student Attitudes; Student Evaluation

ABSTRACT

The general purpose of this study was to compare the differences in middle school students' mathematics achievement, their changes in attitude towards mathematics, and their attitude towards evaluation when evaluated with two different measurement strategies. The primary purpose of the study was to compare aspects of criterion-referenced and norm-referenced evaluation within selected sixth and seventh grade mathematics classes at the University of Northern Colorado Laboratory School. The design for this investigation was quasi-experimental nonequivalent control group design. Ninety-five students were assessed in regard to mathematics achievement, attitude towards mathematics, and attitude towards elevation at the beginning and again at the termination of the 12-week trimester. Overall, students obtained higher achievement scores when evaluated using a criterion-referenced method keyed to the specific performance objectives of an individualized instructional program. Students evaluated by criterion-referenced methods demonstrated significantly more positive attitudes towards the subject than did those middle school students evaluated by norm-referenced methods. Middle school students indicated no preference for either the criterion-referenced or the norm-referenced methods of evaluation. (Author/PN)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED229451

# Do Pupil Grading Methods Effect Middle School Students' Achievement: A Comparison of Criterion-Referenced Versus Norm-Referenced Evaluation

Kenneth T. Wilburn

University of North Florida

Barry C. Felps

Wolfson Senior High School, Jacksonville, Florida

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K.T. Wilburn

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 362

Grading Methods

Abstract

The general purpose of this study was to compare the differences in middle school students' mathematics achievement, their changes in attitude towards mathematics and their attitude towards evaluation when evaluated with two different measurement strategies. The primary purpose of the study was to compare aspects of criterion-referenced and norm-referenced evaluation within selected sixth and seventh grade mathematics classes at the University of Northern Colorado Laboratory School. This study was concerned with three specific questions: (1) Is there a difference in mathematics achievement of students evaluated by criterion-referenced methods and norm-referenced methods? (2) Is there a difference in the attitude of students evaluated by criterion-referenced methods and norm-referenced methods towards mathematics? (3) Is there a difference in the attitude of students evaluated by criterion-referenced methods and norm-referenced methods towards evaluation? The design for this investigation was quasi-experimental non-equivalent control group design. The population for this study was provided by the Laboratory School of the University of Northern Colorado. Those students in the Middle School's sixth and seventh grade class were normally divided into four general mathematics classes by their daily class schedule. Two classes, one sixth grade and one seventh grade, were randomly selected as experimental groups, leaving the two remaining classes as control groups. The purpose of these classes was to explore general mathematics topics. Prior to the experiment the researcher and the participating teacher developed specific performance objectives so designed as to outline three four-week instructional units for all groups. The content of both the control and experimental groups was the same for all similar groups. All four classes

were taught by the same instructor. Three instruments were used to generate pretest and posttest scores for comparison. The researcher and the participating teacher developed criterion-referenced tests of mathematics achievement keyed to the instructional performance objectives to measure the participants' mathematical progress. The researcher developed a Likert-type (equal-appearing interval scale) attitude sale to register the extent of the participants' agreement or disagreement with a set of predetermined basic concepts concerning criterion-referenced and norm-referenced evaluation. The researchers also developed selected a Likert-type attitude instrument in order to measure the participants' attitude towards mathematics. Participants included ninety-five sixth and seventh grade middle school students at the University of Northern Colorado Laboratory School during the spring tri-semester of 1975. All participants were assessed in regard to mathematics achievement, attitude towards mathematics and attitude towards elevation at the beginning and again at the termination of the twelve-week tri-semester. The findings in this experiment led to the following conclusion: Overall, the achievement of middle school students is significantly affected by the method used for evaluation in the instructional process. (2) Higher achieving middle school students are less affected by the method evaluation used in the instructional process than are lower achieving students. (3) Lower achieving middle school students are affected more by the method of evaluation used in the instructional process than are higher achieving students. (4) Overall, middle school students obtained higher achievement scores when evaluated using a criterion-referenced method keyed to the specific performance objectives of an individualized instructional program.

(5) A middle school student's attitude towards a subject or course of study is significantly affected by the type of evaluation system used. (6) Overall, middle school students evaluated by criterion-referenced methods demonstrated significantly more positive attitudes towards the subject than did those middle school students evaluated by norm-referenced methods. (7) Middle school students indicated no preference for either the criterion-referenced or the norm-referenced methods of evaluation.

## Introduction

The purpose of this study was to collect, organize and interpret the effects of pupil evaluation (grading) upon students. Specifically, the study was concerned with three questions: (1) What is the effect of pupil evaluation on pupil achievement? (2) What is the relationship between pupil evaluation and the attitude of student towards the subject under study? (3) Do students have a bias towards or a preference for a particular type of pupil evaluation method?

The publication of Wad-ja-get? by Simon and Napier (1971) marked the beginning of a new era of professional debate about an old nemesis--grading. While Mager (1962) called for specific instructional objectives, Popham (1971) revealed the merits of criterion-referenced tests and Glasser (1969) toured the country advocating "schools without failure"; most teachers were trying to determine if the student who only achieved two of ten behavioral objectives on his criterion-referenced tests would have his self-concept destroyed and become another victim of "dehumanized education" if he got an "F" in math! The terminology may have changed, but the "grading game" was still being played in most American schools.

However, the "grading game" was suffering from a persistent problem. The primary problem being the elementary and secondary school teachers' misuse of the bell-curve and its applications to grading practices. Bloom (1969) stated:

> We have for so long used the normal curve in grading students that we have come to believe in it. Our achievement measures are designed to detect differences among our

learners, even if the differences are trivial in terms of subject matter....

In any group of students we expect to have some small percent receive an "A" grade. We are surprised when the percentage differs greatly from about ten percent. We are also prepared to fail and equal proportion of students. Quite frequently, this failure is determined by rank order of the students in the groups rather than by their failure to grasp the essential ideas of the course.

## Statement of the Problem

Do teacher grading methods effect pupil achievement? For generations we have accepted the following three stage instructional model.

---

Insert Figure I about here

---

The fact that we believe in a high correlation between teaching and learning is supported by the overwhelming volume of research on teaching media and methodology. A consistent body of research exists concerning the ways in which pupils learn. But what is the effect of the evaluation process on pupil achievement? Is pupil evaluation an unrelated assessment process or is it an integral part of the teaching, learning model that may not assess instruction but have effect upon the success of that instruction?

This research activity was based on the concept that pupil evaluation

is an integral part of the teaching-learning process and therefore may affect pupil achievement. In addition, all other teaching-learning activities such as methods, materials and other student related activities must be planned with respect to the pupil evaluation methods used.

## Description of Procedures and Design

This experiment was conducted at the University of Northern Colorado Laboratory School at Greeley, Colorado. The school maintains a K-12 program as a department of the College of Education. It is designed to provide a facility for research and experimentation with new teaching methods and to offer preteaching experiences for the college's professional teacher education programs. Approximately 600 students are enrolled on a first-come, first-served basis. The population selected for this study was the 97 students enrolled in the middle school's sixth and seventh grade mathematics classes. These classes were designed to explore topics of general mathematics and the subject matter content of the courses was structured so that all similar student groups were exposed to the same established curriculum.

At the beginning fo the 1974-75 school year the sixth and seventh grade students were randomly assigned to an A.M. or P.M. class, and all four classes were taught by the same instructor. The data for their experiment was collected during the three twelve-week trimesters.

---

Insert Figure II about here

---

The mathematics classes of the middle school program were divided into two sections for each grade by their normal school schedule, thus providing four sections for the experiment. One section of each grade was randomly chosen by the toss of a coin to be the experimental group while the other group remained as the control group. The experimental group was to be evaluated by using a criterion-referenced evaluation method while the control group used a norm-referenced method.

Prior to the experiment the participants were administered three pretest instruments. First, a Mathematics Achievement Test, designed by the researchers containing randomly selected test items matched with the performance objectives developed for the experiment was given to each of the four classes. Secondly, all participants were administered a Linkert-type Grading Attitude Questionnaire containing items designed by the researchers to determine if the participants indicated any preference for norm-referenced or criterion evaluative methods. Finally, all participants were administered a student Attitude-Questionnaire developed for use with middle school students by the researchers to assess the participants attitude toward mathematics. An analysis of pretest data indicated no significant difference among the four mathematics groups. At the conclusion of the pretest period all groups began a series of four instructional units. The objectives of the units were designed so that all similar groups covered the same content. The instructor for all groups was the same, and every attempt was made to ensure that the only controlable difference between the two groups was the method of pupil evaluation.

## Experimental Treatment

The experimental group (criterion-referenced group) was evaluated at the conclusion of each instructional unit (approximately every four weeks) using an instructional unit criterion-referenced mathematics test developed by the researchers that was keyed to the specific performance objectives of the instructional unit. Each student's grade was determined by comparing his/her score to a predetermined performance standard. In addition, recognition was given for exceeding the basic performance standard or basic skill level. For purposes of continuity with the control group letter grades were awarded to the experimental group using a letter code developed previously by Emmert and Wilburn (1974). Basically, this code was as follows:

---

Insert Grading Code Here

---

## Control Group

The control group (norm-referenced) was evaluated at the conclusion of each instructional unit using an instructional unit achievement test identical to the experimental group. However, each student's grade was not determined by comparing his/her score to a predetermined performance standard, but rather to each member of the group. Grades were given so as to outline a normal distribution. The traditional A,B,C, letter code was used in accordance with the distribution as outlined in Figure 3.

Insert Figure III about here

The experiment continued in this manner through four instructional units over a period of twelve weeks. At the conclusion of the experiment all participants were administered the Mathematics Achievement Posttest, Grading Attitude Questionnaire and Student Attitude Questionnaire.

## Statistical Analysis

The statistical model for this experiment utilized a 2 x 4 factorial design. The students were categorized by pretest achievement test scores and sex. Three analyses were made; one for each independent variable (achievement, methematics attitude, and attitude toward grading) by comparing means of the groups on pretest and posttest scores. This statistical model is outlined in Figure IV.

Insert Figure IV about here

The statistical procedure utilized in this process was factoral analysis of variance as conducted in the Biomedical 05V analysis of variance to look at differences in mean pretest and posttest scores in regard to each independent variable and each possible interaction factor. The final statistical test for significance was the two tailed F test at the .05 significance level.

## Analysis and Summary of the Data

### Mathematics Achievement

Mathematics achievement scores were obtained from the Mathematics Achievement Test administered before and after the experiment. The result of the analysis of variance of the mathematics achievement gain scores with grade as covariance appears in Table 1.

---

Insert Table 1 about here

---

The analysis of variance of the mathematics achievement gain scores produced significant results concerning pretest level effect, achievement by grade interaction, and level by sex interaction. The differences between means of gain scores of the pre-test level effect are presented in Table 2.

---

Insert Table 2 about here

---

The results of this analysis indicated that there is a significant difference in achievement between students with low pretest scores depending upon whether they were evaluated by norm-referenced or criterion-referenced methods. This difference is illustrated in Table 3.

---

Insert Table 3 about here

---

Mathematics Attitude

The results of the analysis performed on the data from the pupil Mathematics Attitude Questionnaire revealed that students evaluated by the criterion-referenced methods received significantly higher mathematics attitude scores than those evaluated by the norm-referenced methods.

---

Insert Table 4 about here

---

Table 4 indicates that the type of evaluation method demonstrated a significant relationship with the method of evaluation used. Further examination of the data also indicated that in addition to the overll significant effect of criterion-referenced evaluation methods on the experimental groups, the grade of the participants produced a significant interaction. This interaction illustrates that the seventh grade participants received significantly higher mathematics attitude scores than their sixth grade counter parts.

Attitude Toward Evaluation

The date provided by the QPAO revealed no statistical significant differences for either criterion-referenced or norm-referenced system. There seemed to be no significant preference participants for either evaluation system as indicated by their scores on the pupil grading attitude questionnaire.

## Summary, Conclusions and Implications

The researchers, in this study, attempted to determine whether students working under a "normal curve" approach to grading or students under a criterion-referenced system that allows grades to be distributed according to relative performance and improvement in regard to predetermined performance standards would reveal greater achievement gain scores. The individualized criterion-referenced approach developed in this study was based on criterion-referenced measurement techniques. The students being evaluated by this individualized criterion-referenced system were not restricted by a predetermined curve or normal distribution. This is to say, no predetermined grade quotas were set on grade categories nor were proportions among categories attempted. If there was a predetermined category, it was to have all students achieve the highest possible achievement score. The findings in this experiment led to the following conclusions. 1) Overall, middle school students obtained higher achievement scores when evaluated using a criterion-referenced method. 2) Higher achieving middle school students were less affected by the method of elevation used in the instructional process than were average and lower achieving students. 3) Lower achieving middle school students were affected more by the method of evaluation used in the instructional process than were higher achieving students. 4) Overall, middle school students evaluated by criterion-referenced methods demonstrated significantly more positive attitudes towards the subject under study than did those middle school students evaluated by norm-referenced methods. 5) Middle school students indicated no preference for either the criterion-referenced

norm-referenced methods of evaluation.

Implications

Because the middle school student is undergoing a transient period marked by the youngster's transition from dependence upon the family for security to a similar dependence on the peer group, he/she demands special considerations in selecting evaluative methods and techniques. The evaluation system, to be most effective, should not contribute to negative peer group pressures and should attempt to maximize the potential benefits of the changing atttitudes of the transient learner. The concern of the middle school for the individual and his/her opportunity for individualized learning should be extended to encompass the evaluative methods of the school program.

The implications of this study in regard to the effect of evaluation on achievement may be summarized as follows: (1) The traditional norm-referenced grading system restricts the academic achievement of most middle school students. (2) A performance-based, criterion-referenced, instructional and evaluative strategy significantly increases the academic ability of middle school students when compared to traditional normative methods. (3) Schools must begin to individualize the evaluation segment of the teaching-learnint-evaluation process to best fit the academic grade level and abilities of each group of students and/or individuals. (4) The use of criterion-referenced evaluation system is of particular benefit to low achieving students.

One of the strongest drives of the middle school student is the basic need for an identity, a belief that he/she is someone different from others, that this someone is important to others and that they see

see themselves as worthwhile. Since the middle school attempts to address itself to the need by maintaining an atmosphere of basic respect for individual differences while providing an environment where the opportunity to succeed is insured for all students, we believe that the needs of the middle school students are best served by an evaluation system that is individualized and criterion-referenced. The implications for the middle school clearly seem to be that taking students out of the competitive atmosphere of the traditional norm-referenced pupil evaluation system and placing them in a less competitive and less threatening atmosphere of a performance-based, criterion-referenced pupil evaluation system more appropriately alines the evaluative process with the basic need of middle school students.

Even though we recognize that the probability for any pupil evaluation system being 100 percent effective for every teacher and every child is practically non-existent, the school cannot refuse to be accountable to the student, parent, and the community in attempting to realistically evaluate the educational progress of each student in regard to his/her own abilities.

References

Bloom, Benjamin. Learning for Mastery. Topical Papers and Reprints, 1969, 1, 2. (Evaluation Comment, 1968-69, 2 (1, Pt.2.). Center for the Evaluation of Instructional Programs, UCLA)

Emmert, Barbara & Wilburn, Kenneth T. An Individualized Criterion Based Reporting System. National Association of Laboratory Schools Bulletin, 1974, 15, 1-9.

Glasser, William. Schools Without Failure. New York: Harper and Row, 1969.

Kirschenbaum, Howard. Simon, Sidney B. & Napier, Rodney W. Wad-ja-get? The Grading Game in American Education. New York: Holt Publishing Company, Inc., 1971.

Mager, Robert F. Developing Attitudes Towards Learning. Palo Alto, California: Fearson Publishers.(Lear Siegler, Educational Division, 1961).

Popham, W. James. Criterion Referenced Measurement: An Introduction. New Jersey: Educational Technology Publications, 1971

Figure I

Traditional Instructional Model

Figure II

Experimental Design

| | | PRETEST | TREATMENT | POSTTEST |
|---|---|---|---|---|
| 6th Grade A.M. & 7th Grade P.M. Classes | Experimental Group | $T_{1E}$ | X | $T_{2E}$ |
| 6th Grade P.M. & 7th Grade A.M. | Control Group | $T_{1C}$ | | $T_{2C}$ |

Grading Code

| | | | |
|---|---|---|---|
| U | - | Unsatisfactory | Student did not meet basic level. |
| S | - | Staisfactory | Student did not meet basic level but showed improvement. |
| B.S. | - | Basic Skill | Student met basic performance level (predetermined). |
| A | - | Proficiency | Student met and exceeded basic skill level and correctly answered 90 percent of the items. |
| M | - | Mastery | Student, after meeting basic skill level, contracted with the instructor and completed a special project demonstrating application of the objectives. |

Figure III

Traditional Grade Distribution

| 10% | 20% | 40% | 20% | 10% |
|---|---|---|---|---|
| F | D | C | B | A |

Figure IV

Statistical Model

| | Norm-Referenced | Criterion-Referenced |
|---|---|---|
| High Achievement Pretest Scores | | |
| Male | | |
| Female | | |
| Low Achievement Pretest Scores | | |
| Male | | |
| Female | | |
| | | |

TABLE 1

Analysis of Variance: Mathematics Achievement

| Source of Variance | Degree of Freedom | Sums of Squares | F Ratio |
|---|---|---|---|
| Pretest level Effect | 1 | 484.431 | 15.18** |
| Achievement by Grade Interaction | 1 | 403.685 | 12.65** |
| Level by Sex Interaction | 1 | 126.589 | 3.97** |
| Error Within | 76 | 2425.845 | |

*P $<$ .05, ** P $>$ .01

TABLE 2

Post-Test Means of Gain Scores of Pre-Test Achievement by Level Interaction

| | | Evaluation Type | |
|---|---|---|---|
| | | Mean Norm-Referenced Gain | Mean Criterion-Referenced Gain |
| Pre-test Level | High | 21.99 | 19.34 |
| | Low | 12.25 | 18.95* |

* $P < .05$

TABLE 3

Difference in Pre-test and Post-test Mean Scores of Low Level Students